<u>WHAT IS CLAIMED IS</u>:

1.   A method for representing the structure of a polyketide produced by a modular polyketide synthase, said method comprising the steps of:

5   (a)   defining a set of monomer units of which said polyketide is composed,

             (b)   assigning an alphanumeric symbol or symbols to each different monomer unit in said set,

             (c)   identifying one or more monomers in said set that is present in said
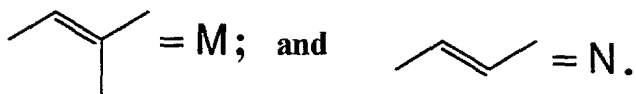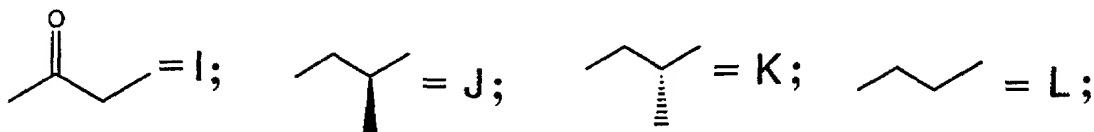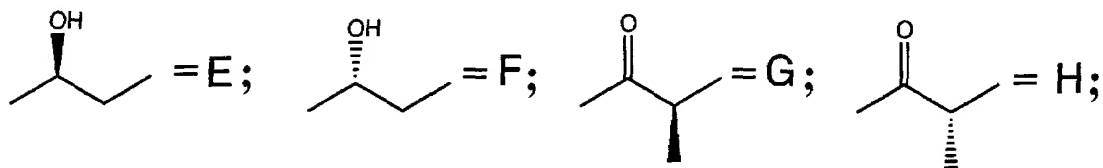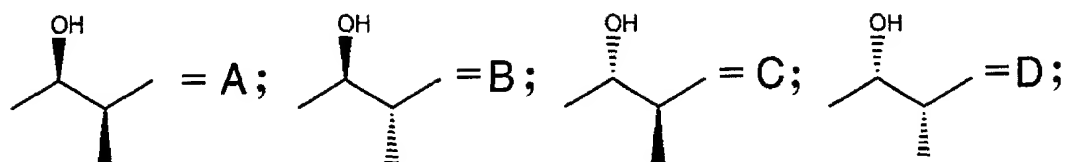
10   polyketide, and

             (d)   composing a string of said symbols ordered in a manner reflecting the order in which said monomers occurs in said polyketide, wherein said string of symbols represents the structure of said polyketide.

15   2.   The method of claim 1, wherein said monomer set comprises two-carbon unit monomers, wherein a first carbon of said unit is substituted with hydrogen or methyl, and a second carbon of said unit is substituted with oxygen, hydroxy, or hydrogen, and said two carbon unit comprises either a single or a double bond between said first and second carbons.

20

3.   The method of claim 2, wherein said monomer set additionally comprises one or more members selected from the group consisting of two carbon unit monomers in which said first carbon is substituted with hydroxy, methoxy, or ethyl; a moiety corresponding to an amino acid or amino acid derivative incorporated into a PKS by a non-ribosomal peptide synthase; a moiety corresponding to a structure incorporated into a polyketide by

25   an AMP ligase or a CoA ligase; and a moiety corresponding to a structure corresponding to a structure in a polyketide after modification by a polyketide modification enzyme.

4. The method of claim 2 wherein the set of monomer unit and corresponding symbol comprises:

OH [structure] = A; OH [structure] = B; OH [structure] = C; OH [structure] = D;

OH [structure] = E; OH [structure] = F; O [structure] = G; O [structure] = H;

O [structure] = I; [structure] = J; [structure] = K; [structure] = L;

[structure] = M; and [structure] = N.

5. The method of claim 4 wherein the set of monomer unit further comprises a miscellaneous monomer that is assigned the symbol Q.

5

6. The method of claim 4 wherein the set of monomer unit and corresponding symbol further comprises
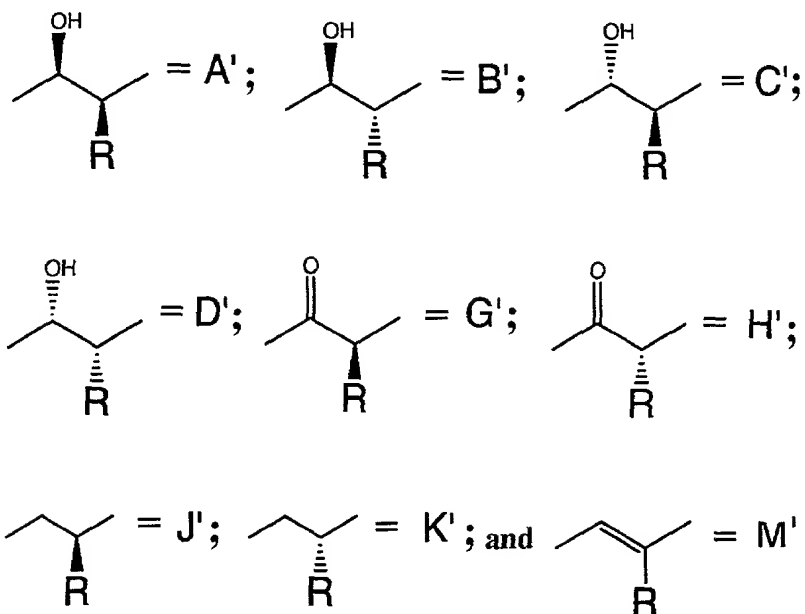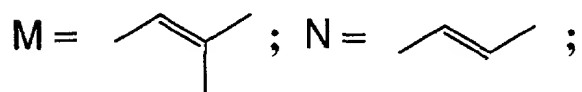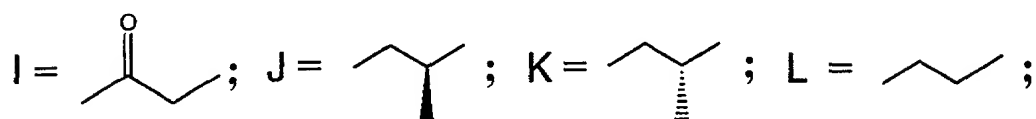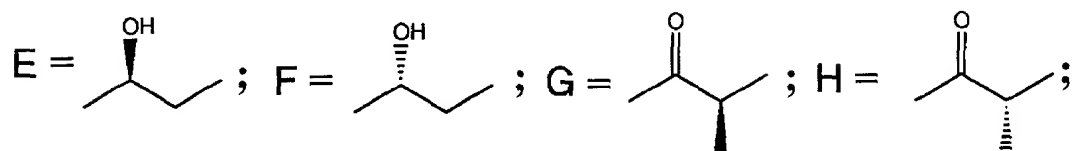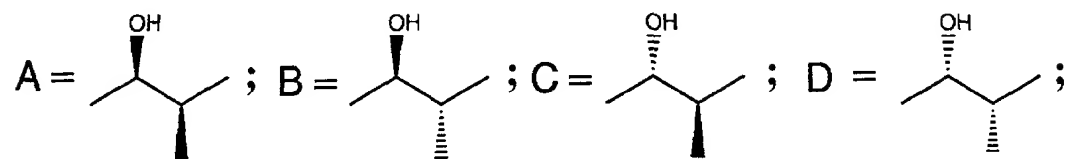

7. A database of polyketides, in which each said member is represented by a string of alpha-numeric symbols, wherein said symbols represent structural subunits of said polyketide, and said string represents the order in which such subunits occur in said polyketide.

8. The database of claim 7 that includes at least 100 different polyketides.

9. The database of claim 7 wherein each said member is represented by a CHUCKLES string.

10. The database of claim 7 wherein each said member is represented by an annotated CHUCKLES string.

43

11.   The database of claim 7 wherein the symbol and its corresponding structural subunit are selected from the group consisting of

A = (structure, OH); B = (structure, OH); C = (structure, OH); D = (structure, OH);

E = (structure, OH); F = (structure, OH); G = (structure, O); H = (structure, O);

I = (structure, O); J = (structure); K = (structure); L = (structure);

M = (structure); N = (structure);

and Q for a miscellaneous monomer.

5

44

12. The database of claim 7 wherein the symbol and its corresponding structural subunit are selected from the group consisting of
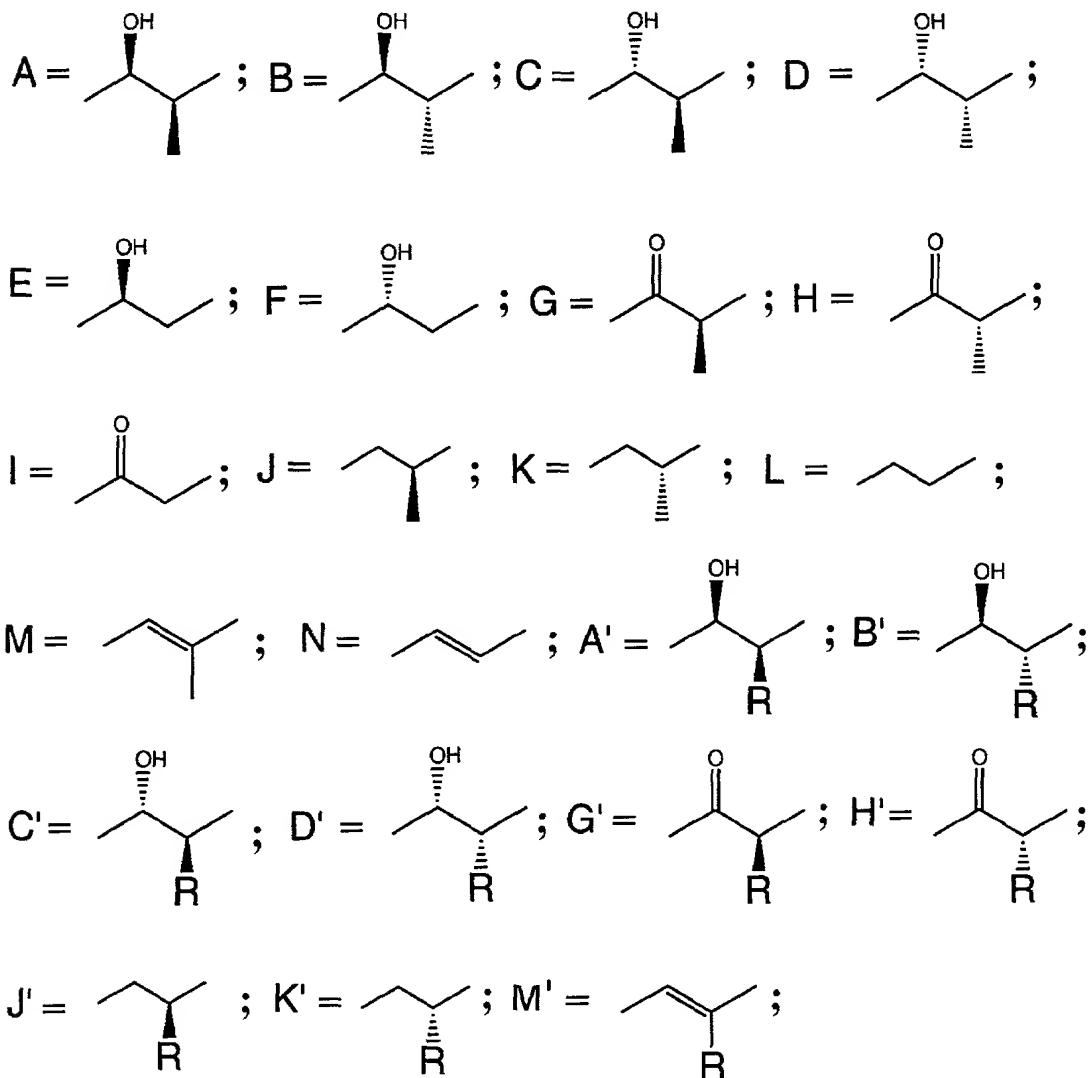
and Q for a miscellaneous monomer.

5

13. A database of polyketides, in which each said member is represented by a linearized representation of said polyketide.

14.   A method of designing a PKS gene capable of producing a desired polyketide, which method comprises:

(a)   defining a string of alphanumeric symbols representing the structure of said polyketide,

5   (b)   comparing said string to a database of strings of alphanumeric symbols representing polyketides produced by PKS genes,

(c)   identifying common elements in said string representing the structure of said polyketide with elements in said strings in said database, and

(d)   generating one or more new strings from elements identified in step (b)

10   that match said string representing the structure of said polyketide, wherein said new string defines a PKS gene capable of producing said polyketide.

15.   The method of claim 14, wherein all possible PKS genes encoding a desired polyketide from said database are generated and displayed.

15

16.   The method of claim 14, wherein said new strings generated in step (d) are rated and displayed in an order based on one or more parameters.

17.   The method of claim 16, wherein said parameters are selected from the group

20   consisting of number of non-native module interfaces and number of non-native protein interfaces.